

## Python Script - Data Cleaning and Insights

```
import pandas as pd

import os


# Step 1: Load the raw XLSX data

file_path = r"C:\Users\hp\Downloads\Raw_superstore_sales.xlsx"

df = pd.read_excel(file_path, engine="openpyxl")


# Step 2: Data Cleaning

df_cleaned = df.drop_duplicates()

numeric_columns = ['sales', 'profit', 'quantity', 'shipping_cost']

categorical_columns = ['region', 'product_category', 'order_priority']

for col in numeric_columns:

    if col in df_cleaned.columns:

        df_cleaned[col] = pd.to_numeric(df_cleaned[col], errors='coerce').fillna(0)


for col in categorical_columns:

    if col in df_cleaned.columns:

        df_cleaned[col] = df_cleaned[col].fillna('Unknown')


date_columns = ['order_date', 'ship_date']

for col in date_columns:

    if col in df_cleaned.columns:

        df_cleaned[col] = pd.to_datetime(df_cleaned[col], errors='coerce')
```

```
if 'sales' in df_cleaned.columns and 'profit' in df_cleaned.columns:
```

```
    df_cleaned['Profit Margin'] = (df_cleaned['profit'] / df_cleaned['sales']).fillna(0)
```

```
text_columns = df_cleaned.select_dtypes(include=['object']).columns
```

```
for col in text_columns:
```

```
    df_cleaned[col] = df_cleaned[col].str.strip().str.title()
```

```
# Step 3: Save the cleaned data to the specified folder
```

```
output_folder = r"C:\Users\hp\OneDrive\Desktop\New folder (4)"
```

```
os.makedirs(output_folder, exist_ok=True)
```

```
df_cleaned.to_excel(os.path.join(output_folder, "cleaned_superstore_sales.xlsx"), index=False)
```

```
df_cleaned.to_csv(os.path.join(output_folder, "cleaned_superstore_sales.csv"), index=False)
```

```
# Step 4: Generate Basic Insights
```

```
total_sales = df_cleaned['sales'].sum()
```

```
total_profit = df_cleaned['profit'].sum()
```

```
average_profit_margin = df_cleaned['Profit Margin'].mean()
```

```
# Step 5: Save Insights to File
```

```
insights = {
```

```
    "Total Sales": total_sales,
```

```
    "Total Profit": total_profit,
```

```
    "Average Profit Margin": average_profit_margin,
```

```
}
```

```
insights_file = os.path.join(output_folder, "cleaning_insights.txt")
```

```
with open(insights_file, "w") as f:
```

```
    for key, value in insights.items():
```

```
        f.write(f"{key}: {value}\n")
```